

Link to Success: How Blogs Build an Audience by Promoting Rivals

June 2006

Dina Mayzlin, Yale School of Management
Hema Yoganarasimhan, Yale School of Management

DRAFT – COMMENTS WELCOME

Please do not cite or circulate without permission

Abstract

Weblogs or blogs are typically websites maintained by a single person or group of persons on a specific topic. Blogs started off as single person diary entries, but now are almost part of the mainstream media. The key research question of this paper is - Why are blogs so popular given that blogs could be potentially biased or contain little informational content. The unregulated nature of the blogosphere raises the question of how the blogs' quality can be maintained.

We seek to explain the popularity and quality content of blogs by the phenomenon which we call "monitoring through linking". Blogs typically have outgoing links to other blogs which are created by the blog's author. So, if the blog author (*blogger*) considers some other blog to be good, he links to that particular blog. This is both interesting and intriguing since linking to another blog is an advertisement for that blog.

This differs from traditional consumer markets where i) competing firms don't promote each other; ii) competing firms don't monitor each others quality. Effectively, bloggers both promote and compete with each other. Thus, the blogs which succeed are the ones which pass the scrutiny of the blogging community.

We present a model where blogs are differentiated along two dimensions: the ability to produce original content and the ability to link to other blogs with original content. Thus, by linking a blog demonstrates its ability to link to original content in future periods. The downside of a link is that it is a recommendation of a competing blog: it demonstrates that the other blog is more likely to deliver original content next period. We show that linking is an equilibrium strategy outcome under certain conditions. We also examine the consequences of this monitoring system in a variety of contexts.

1. Introduction

By some estimates, about 35 million workers visit blogs and on average spend about 3.5 hours of the work week reading or posting comments in blogs.¹ On May 12, 2006, Technorati (a web log or “blog” search engine), was tracking 39.1 million sites. On that same day, a visitor to the blog daddytypes.com (“the weblog for new dads”) could find posts in several categories. For example, there was a post about zwangerschapssnorkel – a plunger-like device applied to the mother’s belly button that ostensibly improves the flow of oxygen to the fetus. There was also a picture featuring an “unidentified Red Sox fan” pushing a baby in a Gecko stroller. The author of the blog went on to note, “on the rare occasions when I see a Gecko in the wild, its clean, simple lines and understated, old-school color scheme really stands apart from the pack.” Finally, there was a post announcing a two-day sale at Netto Collection, an upscale kid furniture store in NYC. One particularly notable aspect about the last post was that it credited (and linked to) another blog, daddydrama.com (“a blog that’s all about the baby (well, hip dads and chic moms too)!”), with the original post on the sale. In fact, while daddytypes posted the information on May 11 (the first day of the sale), daddydrama had the information posted on May 9.

The example above illustrates several interesting aspects of blog sites. The first is that the author of the blog is an amateur in the sense that he or she is not employed by a media company.³ The second is that the blog provides potentially useful information to its readers. For example, daddytypes provides extensive stroller recommendations. The third is that there

¹ Johnson, “What Blogs Cost American Business,” (October 24, 2005), *Advertising Age*.

³ The NYTimes featured the author of daddytypes, Greg Allen, in an article that described him as a filmmaker and a writer. (“Changin’ in the Boys’ Room,” Newman, Feb 5, 2006)

is a large number of blogs for Dads – daddytypes alone lists 18 “daddy blogs” on its blog roll (a list of blogs usually featured on the right hand side of the site). Given constraints on one’s time, how can a reader find a blog that would supply him with reliable and useful information? Finally, given these constraints, why would an author link to another blog in the same category, thus possibly jeopardizing losing a reader to another site?

The phenomenon of linking to (or recommending) a rival is not confined to this space. For example, on its book pages Amazon features links to other sellers who have a new or a used version of the book, often at a discounted price. While Amazon takes a commission from any sales that is generated when the customer clicks on the link, it is nonetheless surprising that Amazon would facilitate building a relationship between its customer and another bookstore. While we believe that recommending rivals can happen in a number of contexts, we focus on the blogosphere (a term used to describe the universe of web logs) because it represents a particularly compelling example of this phenomenon. In addition, we believe that this incentive to link is at least partly responsible for the phenomenal success of the blogosphere.

Why would a blogger link to a rival blog? We suggest that bloggers differ along two dimensions: 1) the ability to produce original information, and 2) the ability to find information in other sites. Using the example above, by linking to daddydrama’s announcement of the furniture sale, Greg Allen signals to the reader that he is able to not only provide original content (a discussion on strollers, etc) but also that he is able to provide a link to information produced by other sites. Similarly, we suggest that one reason behind Amazon’s decision to create Amazon marketplace which connects consumers to

other booksellers is that this allows Amazon to be viewed as a “destination” site: the consumer will go to Amazon next time because she will be able to find a book either on Amazon or on another site through Amazon.

Note that as a by-product of the set-up above, there exists a self-sustaining system of quality monitoring between different blogs. That is, by linking to a blog with information, a consumer is able to find information by following links and a blogger who is more likely to produce information has a higher readership. This effect is further accentuated by search engines that commonly give higher placement to sites with more incoming links. There have been a number of articles in computer science that argued that links are a good metric of web site quality. This paper, on the other hand, presents a micro model why linking is an equilibrium strategy.

The rest of the paper is organized as follows: in Section 2, we discuss previous literature, in Section 3, we present a model of the incentive to link, and in Section 4 we conclude.

2. Previous Work

The first stream of literature that relates to our paper deals with the role of hyperlinks in the Internet. This has been studied in computer science since Kleinberg’s seminal 1999 paper. Kleinberg proposed that links have valuable information because they reflect the subjective judgment of the author who made them. The paper was written when most search engines were text based and the results they produced were dismal. The hyperlink based algorithm

developed by Kleinberg is still the basis of all the search engines that we see today. In his famous example, he uses the case of someone searching for *automakers* (Toyota, Nissan etc). If the person were to use a text based search engine then it is unlikely that she will ever get to the websites of auto manufacturers since they never use the *automaker* on their website. Instead a search engine using links could locate the websites easily. The Kleinberg algorithm essentially captures the idea that people who make websites are intelligent beings and the websites they link to are ones with information. The collective wisdom of the mob separates the good from the bad.

When Kleinberg proposed his algorithm for search engines, he assumed that the world was divided into hubs and authorities. Authorities were homepages or authoritative pages on some subject, while hubs were web pages which linked to authorities. Brin and Page XX furthered the idea of link-based search by eliminating the hubs-and-authorities structure and introducing PageRank – an algorithm which computes the page rank of a webpage for a given query. The internet isn't divided into hubs and authorities and from a practical perspective PageRank is a viable algorithm and is actually the basic framework behind Google.

While the computer science literature has assumed that there is valuable information in links, our aim here is to explain the existence of these networks by modeling the links (incentive structure) of the bloggers. Thus, while the work in computer science takes the network as given and tries to retrieve information from the network, we model the incentives of the potential players (of a network) using links to derive the network itself.

While most of the literature has assumed network configuration as given, one notable exception is work by Bala & Goyal and Jackson et al which studies how networks are formed. Though we too study incentives in link-formation, there are some key differences in our approaches. First of all, the conceptualization of a link in Bala and Goyal is that of a friendship link. That is, once a link is formed, information can flow through the link. For example, if I have lunch with a colleague once a week, I may learn about her research. In the blogosphere, of course, the flow of information does not depend on the link structure in the sense that one blogger can find out about the information on another blog simply by reading it.

In addition, while in Bala and Goyal the game is essentially between the nodes only, we introduce another player – the reader. Hence, in our paper there exists the potential for a posted link to lead the reader to another blog. Thus, while in Bala & Goyal, each link is costly to the initiator, here creating the links is costless, and the potential downside is due to the signaling value only. In contrast to past research, we recognize the fact that links are not always beneficial to the linking blog under the presence of a third party who is choosing between the different sites.

The third relevant stream of literature deals with mutual monitoring in teams. For example, Bowles and Gintis (1998) model mutual monitoring in teams when the members are residual claimants. Varian (1990) models how incentive mechanisms can be designed in such a way that agents would monitor each other. Studies of Grameen Bank in Bangladesh (**cite) have shown that social norms are monitors for not defaulting on group loans while people default heavily on individual loans. Past research has dealt with monitoring in principal agent

problems where the principal devices an incentive structure such that agents monitor each others effort. Our model (problem) is essentially different because we are not studying a mechanism design. Here, the agents promote rivals for their own benefit and not to monitor the quality of the blogosphere. Quality monitoring is a by-product of the selfish behavior by the blogs. Though both our model and the previous models in economics deal alike with quality monitoring, the context and the results are very different.

3. Model

3.1 Set-Up

We have a finite game with M risk-neutral consumers and N blogs. For expositional purposes we break up the two periods into sub-periods. We assume that bloggers obtain utility from the size of the readership. This is either due to commercial or social reasons. That is, a blogger with a bigger audience will make more revenue if she chooses to place ads on her site through a service such as Google's Ad Sense. However, perhaps an even bigger motivating force in this context is the social utility that a blogger derives from having a bigger audience.

We model bloggers as producers of information and visitors as consumers of information. We assume that at $t = 1$ a consumer visits a blog at random (let's call it blog A). At $t = 2$, a piece of valuable information becomes available to some bloggers. For example, a blogger may find out the name of the jurist that will be nominated to the Supreme Court or she may find which store will have a sale on summer swimwear. The type of information modeled here is what in the newspaper business is called a "*scoop*". We assume that a blogger must be truth-telling in its posting of a news story. That is, a blogger who simply manufactures a story or makes up a sale will be severely punished by other bloggers and the audience.

Bloggers are differentiated along two dimensions: i) The ability to obtain the scoop and hence to break the news on its site; and ii) The ability to find other blogs that contain information and hence the ability to link to news-breaking blogs. These abilities are assumed to be independent and driven by two different costs: while some bloggers have access to valuable information, others have low search costs when it comes to searching for information in other sites. This gives rise to four types of bloggers characterized by the ability to either post information early on its own site or to post a link (which involves a delay) to another blog which contains information. Specifically, a blog can be either a High (H) type (probability v of acquiring information) or Low (L) type (probability $w < v$) when it comes to breaking the news on its blog and H type (probability p) or L type (probability $q < p$) when it comes to finding information in other blogs (see Table 1). We will refer to the former ability as ability to *break the news* and the latter as the ability to *find the news*. We assume that at the beginning of the game the blogger's type is not observable to any of the game's players (including the blogger himself): there is symmetric uncertainty. As the game goes on, however, the blog reader learns more about the blogger's abilities. The prior probability that blogger is H type on breaking the news is γ , and the prior on H type on finding the news is δ . Thus, the prior probability that a blog breaks the news is $\alpha_0 = \gamma v + (1 - \gamma)w$ and the prior probability that a blog finds the news in other blogs is $\beta_0 = \delta p + (1 - \delta)q$.

Table 1 – Blogger Types

	<i>Probability p of finding info in other blogs.</i>	<i>Probability q of finding info in other blogs.</i>

<i>Prob v of obtaining info.</i>	L,L	L,H
<i>Prob w of obtaining info.</i>	H,L	H,H

We assume that all blogs post the news if they obtain the scoop since there is no strategic reason for a blog to withhold information. If the consumer sees a blog A posting a news-breaking story at $t = 2$, she derives utility u from the post and updates her prior on A's ability to break the news upwards. Otherwise, she derives no utility and updates the prior on A's ability to break the news downwards.

At $t = 3$, bloggers search others' blogs for information. At this time, those bloggers that found blogs with news may simultaneously post links. We require that credit must be given to the blog that originally posted the information. That is, the blog is not allowed to simply appropriate the information without a link. This assumption is reasonably realistic. First, we do observe examples of attribution. For example, the Web site Smoking Gun was almost universally given credit in the mainstream media as well as in the blogosphere for exposing James Frey's memoir "A Million Little Pieces" as largely fictional.⁶ Second, a blog that consistently plagiarizes other's posts may be exposed by other bloggers.⁷ Finally, we assume that a blogger may not be able to manufacture links for the same reasons that he is not able to manufacture new-breaking news.

⁶ <http://www.thesmokinggun.com/jamesfrey/0104061jamesfrey1.html>

⁷ For example, one irate blogger complained in his blog about a discovery that he had been plagiarized, "I recently lost out on a boatload of potential new readers because a blogger plagiarized my work verbatim. A high-traffic blogger ([Michelle Malkin](#)) then unwittingly linked to the plagiarist's blog instead of mine, and I missed out on all the traffic that came before I found the mistake and asked Michelle to fix the link." <http://davidm.blogspot.com/2004/10/serial-plagiarism-exposed.html>

A consumer who has not yet seen the news derives a utility $u - c$ if she sees a link to a news-breaking blog (let's call it blog B). Note that we assume that the value of information declines with time.⁸ For example, if the information is the name of the store that is having a sale, there is urgency in finding out the information since the store may sell out of the consumer's sizes. A consumer who has already seen the news derives no extra utility from the link to blog B.

After seeing a link to site B, the consumer updates upwards her belief on B's ability to break the news. At the same time, depending on the presence of a link and on the equilibrium of the game, the consumer may update her belief on A's ability to find the news. This is the key tension in the model: by linking blogger A may be able to make itself look good, but at the same time he also makes blogger B, a potential competitor in the next period, look good as well.

In order to simplify the analysis, we assume that the consumer does not update her prior on B's ability to find news in other blogs. This is due either to the fact that the consumer simply does not observe B's links (the information from B may be consumed through A's post) or it may be due to the fact that by the time the consumer visits blog B and observes its links, the news has become stale and the links have no signaling value. The results of our analysis are qualitatively the same if we assume that the consumer is able to resolve her uncertainty about B's ability to find the news but the analysis is somewhat more cumbersome. We also assume that the consumer does not learn about the abilities of any

⁸ An alternative way to think of c is as the travel cost that a consumer must pay in order to travel to blog B.

other blog during this time period. Again, this can be either due to constraints on the reader's time or simply because information quickly becomes stale in this environment.

At $t = 4$, the consumer decides which blog she should visit next. If blog A had linked to blog B, she chooses between A and B. If on the other hand, blog A had not linked to another blog, the reader must choose between A and a random blog (let's call it blog C). When making this choice, suppose that the consumer also experiences a reader-site-specific random shock to her perceived utility from that site. In the absence of strategic considerations (which is the case at the last rounds of the game), we assume that all blogs link if they find news.

The perceived utility from a site i at $t = 4$ is a function of whether the blog will either post a news-breaking item or will link to a news-breaking blog. Hence,

$$EU_i = \alpha_i v u + (1 - \alpha_i) \beta_i (u - c) + \varepsilon_i \quad (1)$$

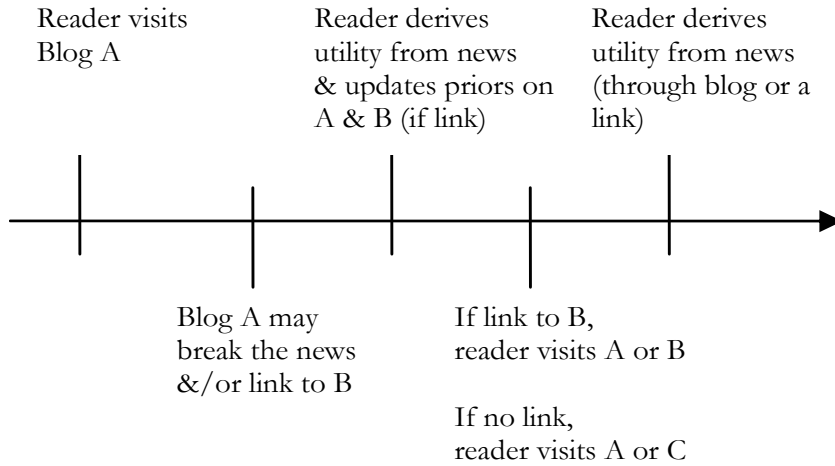
where α_i and β_i are the updated probabilities that the blog will either break the news or find the news respectively. We assume that ε is distributed with cumulative distribution function F .

$$\text{Prob}(i \text{ will be chosen over blog } j) = \text{Prob}(\varepsilon_j - \varepsilon_i < EU_i - EU_j) \quad (2)$$

At $t = 5$, a piece of information is released to some blogs, and blogs who obtain this information go on to post it. If a news item is posted, the reader derives a utility u from the item. At $t = 6$, blogs will link to news-breaking sites if they find a blog that contains the news. If a reader is exposed to the item for the first time, she experiences a utility $u - c$ from the

news item. The game ends after this. The condensed timeline is summarized in Figure 1 below:

Figure 1: Condensed Time Line



3.2 Perfect Bayesian Nash Equilibrium

Here we look for a Perfect Bayesian Nash Equilibrium with respect to the linking behavior of a blogger in the beginning of the game ($t = 3$). At this time, the blogger has either broken the news or has failed to do so and has also either found the news in another blog or has failed to do so. The strategy space of a blogger is her decision to link or not link in each of the four scenarios. This implies that there are 4 possible equilibria (see Table 2). We will refer to the first equilibrium as (Link, Link) where the first cell refers to the action when the site has broken the news and the second one refers to the action when the site did not break the news.

Table 2 – Possible Equilibria

	<i>Broke the news</i>	<i>Did not break the news</i>
1	Link	Link

2	Link	Don't Link
3	Don't Link	Link
4	Don't Link	Don't Link

Each equilibrium generates a different set of posterior beliefs by the consumer. For example, compare the difference in inference generated by no link in (Link, Link) equilibrium versus (Don't Link, Don't Link) equilibrium. In the first equilibrium, a lack of a link is seen as a negative signal on the blog's ability to find the news since in equilibrium the blogs that find the news in other blogs post the link. On the other hand, in (Don't Link, Don't Link) equilibrium blogs choose not to link even when they find the news in other blogs. Hence, no link is not seen as a negative signal and the reader does not change her prior on the blog's ability to find the news.

In addition, as is the case for the majority of signaling models, some actions are not in equilibrium, and hence we have to specify off-equilibrium beliefs. Specifically, if linking is observed but is not played in equilibrium, Bayes' Rule does not apply. (Note that an absence of a link can always be attributed to the fact that the blogger did not find the news and thus is never seen as an out-of-equilibrium action). We assume that a reader updates upwards her beliefs on the blogger's ability to find information when she observes a link, even if linking is not in equilibrium. Note that this is the most reasonable belief, given the structure of the model. That is, a link implies that the blogger found information in other blogs, which makes it more likely that he is the H-type on that dimension. The fact that the blogger plays an out-of-equilibrium strategy should not interfere with that inference

The posterior beliefs are summarized in Table 3. As noted earlier, if blog A contains no link to blog B, the consumer does not learn about the abilities of blog B and next period will choose between A and a random blog C.

Table 3 – Posteriors

Cases	α_A	β_A	α_B	β_B
A breaks the news A links to B	↑	↑	↑	Prior
A breaks the news No link	↑	↓ if (L,L) or (L,DL) prior otherwise	-	-
A fails to break the news Link	↓	↑	↑	Prior
A fails to break the news No Link	↓	↓ if (L,L) or (DL,L) prior otherwise	-	-

As can be seen in Table 3, the potential benefit of linking to another blog is that it credibly signals to the reader that the blogger can find news in other sites. The potential downfall is that the link is a recommendation that a rival is a news-breaking site. Below we analyze when the potential benefit outweighs the costs. That is, when would we expect each equilibrium to hold? How is linking affected by the state of information: does a blogger who already obtained a news-breaking story have a higher or a lower incentive to link to another blog?

We start solving for the equilibrium by analyzing the reader's problem when she chooses a second blog to visit at $t = 4$. The probability that the consumer will choose blog A over another blog j is obtained using (2):

$$\text{Prob}(A \text{ will be chosen over blog } j) = \text{Prob}(\varepsilon_j - \varepsilon_A < EU_A - EU_j) \quad (3)$$

Thus, in the earlier period the blogger's decision to link will be determined by how linking affects the *difference* in expected utilities between blog A and its primary rival. Note also that the primary rival is also endogenous to the linking decision. That is, if blogger A links to another blog B, that news-breaking blog becomes its primary rival. If, however, A does not link to another blog, its primary rival is a random blog C, which will be perceived by the consumer as having an average ability along both dimensions.

We also introduce some notation that will simplify expressions that follow. Let $V(u,d)$ stand for the expected utility that a reader gets from a blog with an updated **u**pwards prior on the ability to break news and an updated **d**ownwards prior on the ability to find news in another blog. (We similarly define $V(d,u)$, $V(d,d)$, and $V(u,u)$).

$$V(u,d) \equiv EU(\alpha_U, \beta_D) = \alpha_U u + (1 - \alpha_U) \beta_D (u - c). \quad (4)$$

If no learning takes place (prior is not updated), we denote that by **n**eutral (**n**). That is,

$$V(u,n) \equiv EU(\alpha_U, \beta_0) = \alpha_U u + (1 - \alpha_U) \beta_0 (u - c). \quad (5)$$

To derive a possible equilibrium, we assume that the reader beliefs are consistent with a given equilibrium, and derive when the blogger chooses a particular equilibrium strategy.

- 1) Assume that the reader beliefs are consistent with (L,L) equilibrium. The sufficient condition for this equilibrium to hold is:

$$V(u,u) - V(u,n) > V(u,d) - V(n,n) \quad (6)$$

- 2) In (L, DL) equilibrium, the blogger chooses to link only if he had broken the news:

⁹ Given the beliefs in this equilibrium, choosing to link if the blogger had broken the news implies that the blogger would link if he had not broken the news.

$$V(u, u) - V(u, n) > V(u, d) - V(n, n) \quad (6)$$

$$V(d, u) - V(u, n) \leq V(d, n) - V(n, n) \quad (7)$$

3) In (DL, L) equilibrium, the blogger chooses to link only if he had not broken the news:

$$V(u, u) - V(u, n) \leq V(u, n) - V(n, n) \quad (8)$$

$$V(d, u) - V(u, n) > V(d, d) - V(n, n) \quad (9)$$

4) For (DL, DL) equilibrium, the blogger never chooses to link:

$$V(d, u) - V(u, n) \leq V(d, n) - V(n, n) \quad (7)$$

Note that holding beliefs constant, a blogger has less incentive to link when he has broken the news. This is due to the fact that the marginal benefit from being perceived as more likely to be H-type on finding the news is lower since he is now perceived as being more likely to be H-type on breaking the news. This is of course due to the assumption that the information that can be found on the two sites is substitutable. However, an added difficulty comes from the change in beliefs across the two states of information. For example, under the (L,DL) equilibrium, the reader's inference is not very punishing following a no link. These two forces may go in opposite directions. In fact, the different beliefs across different equilibria account for the fact that a priori we would expect there to be multiple equilibria in at least some regions.

Next, we describe what equilibria that exist in the different regions of the parameter space.

In the Appendix we show that all the sub-regions described below are not empty:

Proposition 1

¹⁰ Given the beliefs in this equilibrium, choosing not to link of blogger had not broken the news implies that the blogger would not link if he had broken the news.

The space can be broken up into 6 non-intersecting regions below, each with a different set of equilibria.

Region I: (6) holds, (7) does not hold, (8) holds: (DL, L) & (L,L) exist.

Region II: (6) holds, (7) does not hold, (8) does not hold: (L,L).

Region III: (6) does not hold, (7) holds, (9) holds: (DL,DL) & (DL, L).

Region IV: (6) does not hold, (7) holds, (9) does not hold: (DL,DL).

Region V: (6) does not hold, (7) does not hold: (DL,L).

Region VI: (6) holds, (7) holds: all four equilibria exist.

For the proof, see the Appendix.

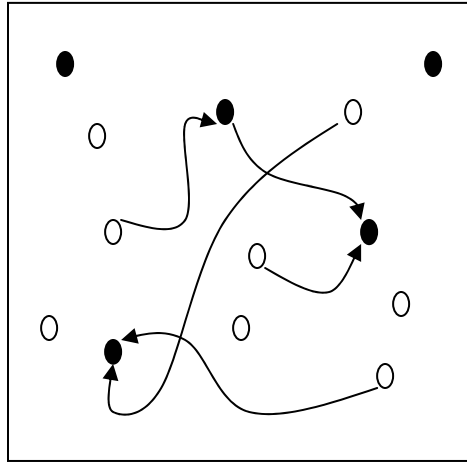
Per our discussion above, (L,DL) equilibrium is never unique. That is, if the belief is such that linking is not expected, and the blogger chooses not to link if he had not broken the news, we would expect the blogger not to link under the same beliefs if he had broken the news. Similarly, if linking is expected, and the blogger chooses to link if he had broken the news, we would expect him to link under the same belief if he had not broken the news. On the other hand, all other equilibria may be unique in certain regions.

Note that one implication of a linking equilibrium is that a blog will be more likely to link to H-type on news-breaking ability since these types of blogs will be more likely to generate a news-breaking piece of information (see Figure 2 below). Thus, this model provides a micro-foundation for why better (more likely to generate a news-breaking story) blogs will have more incoming links. From a blogger's perspective, links give them higher incoming links signify higher readership, higher search engine ranking (which will also lead to

increased readership), and higher recognition and income. Hence, the desire to signal generates an incentive for sites to engage in mutual monitoring: to promote the better sites.

Figure 2 Snapshot of Blogosphere

- Blogs with information; ○ Blogs without information; Arrows – Links.



Next, we investigate how the incentive to link is affected by the changes in parameters of the model. Intuitively, we would expect that the incentive to link would increase if we increase the value of information from learning about the linking ability of a blog. That is, if there is a significant change in priors on linking ability following a link, a blogger would be more likely to link. On the other hand, if there is a significant change in priors following a news-breaking story, a link would significantly help blog B, a potential rival. This would imply that a blog would have less of an incentive to link. This intuition is borne out in the propositions below:

We increase the difference in the abilities to post news-breaking new between the two types in a mean-preserving way. That is, let $\gamma = 0.5$, $v' = v + \frac{\epsilon}{2}$ and $w' = w - \frac{\epsilon}{2}$. (This is a mean-

preserving since $\alpha_0' = 0.5(v + \frac{\epsilon}{2}) + 0.5(w - \frac{\epsilon}{2}) = \alpha_0$. We define the incentive to link as the difference (across linking and not linking) in the difference (between A and a potential rival) in utilities. For example, for the (L, L) equilibrium, the incentive to link is defined as $[V(u, u) - V(u, n)] - [V(u, d) - V(n, n)]$.

Proposition 2

Under all equilibria, the incentive to link -- is *lower* with v' and w' than with v and w .

We similarly allow let $\delta = 0.5$, $p' = p + \frac{\epsilon}{2}$ and $q' = q - \frac{\epsilon}{2}$.

Proposition 3

Under all equilibria, the incentive to link -- is *higher* with p' and q' than with p and q .

Proposition 4

Under all equilibria, an increase in c decreases the incentive to link.

The proofs are given in the Appendix.

The last result implies that a high c , which stands for either the loss of information due to a delay or a travel cost decreases the attractiveness of linking to others. Interestingly, in case of high c , the value from a link is very high too – if the consumer has no time to surf, then he would want to absolutely know where the information is. But then bloggers wouldn't link and hence it would be impossible to sift the good from the bad. In short, high search costs would make searching in the blogosphere an impossible task. This in turn would sharply curtail the popularity of blogs bringing us to a simple yet important implication of this result

– quality monitoring in blogs (and hence their popularity) is a direct consequence of the low click costs on the web.

4. Conclusion

Over the past few years blogs have become extremely popular as sources of news, entertainment etc. The blogosphere is a giant information market where bloggers are information providers (firms in a traditional market) and visitors are consumers of this information. This rapidly expanding market is unique in many ways: 1) It is extremely decentralized, 2) The number of players is huge, and 3) It is very dynamic and the churn in the field is high. From a consumer's perspective, these attributes make it extremely difficult to find useful information in the blogosphere. In short, there is no quality control mechanism in place and given the size and dynamism of the market, it would be impossible for any one consumer to monitor quality. But if useful information in the blogosphere were so elusive, the number of people reading blogs wouldn't be so high. This implies that consumers actually do manage to find useful information.

In this paper, we propose that the nature of blogosphere actually creates an incentive for bloggers to link, which in turn facilitates the consumer's search for information. We show the costs and benefits of linking and find when we would expect different equilibria to hold. We also show how these incentives are changed by changes in heterogeneity between types and costs of information decay. We also show that the consumer beliefs play an important part in determining when linking is optimal.

One important application of this paper is that it provides a micro-foundation for the incentive to link. One interesting application of this paper is in providing a micro-foundation for the algorithm behind Google and other search engines.

References

- [1] Bowles, S. and Gintis, H. (1998), "Mutual monitoring in teams: the effects of residual claimancy and reciprocity," *Research in Economics* 98-08-074e, Santa Fe Institute.
- [2] Ghatak, M. (2002), "Exploiting social networks to alleviate credit market failures: On the endogenous selection of peer groups in microfinance programs," in Proceedings of the Conference on Credit, Trust and Calculation, San Diego, November 15-16, 2002.
- [3] Hölmstrom, B. (1982), "Moral Hazard in Teams," *Bell Journal of Economics*, The RAND Corporation, Vol. 13, No. 2, pages 324-340.
- [4] Islam, M.M., Wahid, A.N.M., and Khan, Z.H. (1993), "The Grameen Bank: Peer Monitoring in a Dual Credit Market, In Wahid, A.N.M. ed. *The Grameen Bank: Poverty relief in Bangladesh*. Westview Press.
- [5] Kleinberg, J. (1999), "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, Vol. 46, No. 5, pp. 604-632.
- [6] Larance, L. Y., (1998), "Building Social Capital from the Center: A Village Level Investigation of the Grameen Bank," Grameen Trust Working Paper No 22.
- [7] Varian, H. R., (1990), "Monitoring agents with other agents," *Journal of Institutional and Theoretical Economics*, pp. 153-174.

Appendix

Proposition 1

We first re-write (6) – (9) by substituting the expressions for expected utilities:

$$1) \quad (L,L): \quad (1 - \alpha_U)(\beta_U - \beta_D)(u - c) > (\alpha_U - \alpha_0)[(1 - \beta_0)u + \beta_0c] \quad (6)$$

$$2) \quad (L, DL): \quad (1 - \alpha_U)(\beta_U - \beta_D)(u - c) > (\alpha_U - \alpha_0)[(1 - \beta_0)u + \beta_0c] \quad (6)$$

$$(1 - \alpha_D)(\beta_U - \beta_0)(u - c) \leq (\alpha_U - \alpha_0)[(1 - \beta_0)u + \beta_0c] \quad (7)$$

$$3) \quad (DL, L): \quad (1 - \alpha_U)(\beta_U - \beta_0)(u - c) \leq (\alpha_U - \alpha_0)[(1 - \beta_0)u + \beta_0c] \quad (8)$$

$$(1 - \alpha_D)(\beta_U - \beta_D)(u - c) > (\alpha_U - \alpha_0)[(1 - \beta_0)u + \beta_0c] \quad (9)$$

$$4) \quad (DL, DL): \quad (1 - \alpha_D)(\beta_U - \beta_0)(u - c) \leq (\alpha_U - \alpha_0)[(1 - \beta_0)u + \beta_0c] \quad (7)$$

Given the set-up of the model, we know that $\alpha_U > \alpha_0 > \alpha_D$ and $\beta_U > \beta_0 > \beta_D$. In fact, (L,L) is characterized by (6) & (9), but we can show that (6) \rightarrow (9). Similarly, (DL, DL) is characterized by (7) & (8), but we can show that (7) \rightarrow (8).

We can also show that if (6) does not hold \rightarrow (8) holds and if (7) does not hold \rightarrow (9) holds.

Since there are 4 different constraints categorizing the equilibria above, there theoretically could be up to 16 different regions. However, since the constraints are not independent, the actual number of regions can be shown to be 6 as described in Proposition 1. In addition, we can show that each region is non-empty with a numerical example.

Region I: Let $u=10, c=1, v=0.9, w=0.3, p=0.9, q=0.2, \gamma=\delta=0.5$.

Region II: Let $u=10, c=1, v=0.9, w=0.4, p=0.9, q=0.2, \gamma=\delta=0.5$.

Region III: Let $u=10, c=1, v=0.9, w=0.01, p=0.9, q=0.2, \gamma=\delta=0.5$.

Region IV: Let $u=10, c=1, v=0.9, w=0.2, p=0.9, q=0.8, \gamma=\delta=0.5$.

Region V: Let $u=10, c=1, v=0.9, w=0.2, p=0.9, q=0.2, \gamma=\delta=0.5$.

Region VI: Let $u=10, c=1, v=0.9, w=0.5, p=0.9, q=0.6, \gamma=\delta=0.5$.

Proposition 2

Here,

$$\alpha_0 = 0.5v + 0.5W, \quad \alpha_0' = 0.5v + 0.5W$$

$$\alpha_U = \frac{v^2 + w^2}{v + w}, \alpha_U' = \frac{(v + \frac{\varepsilon}{2})^2 + (w - \frac{\varepsilon}{2})^2}{v + w}$$

$$\alpha_D = \frac{v(1-v) + w(1-w)}{2 - (v + w)}, \alpha_D' = \frac{(v + \frac{\varepsilon}{2})(1 - v - \frac{\varepsilon}{2}) + (w - \frac{\varepsilon}{2})(1 - w + \frac{\varepsilon}{2})}{2 - (v + w)}$$

As we can see from (6) – (9) in proof of Proposition 1, we need to demonstrate that

$$\text{a) } \frac{1 - \alpha_U'}{\alpha_U' - \alpha_0'} < \frac{1 - \alpha_U}{\alpha_U - \alpha_0} \text{ and}$$

$$\text{b) } \frac{1 - \alpha_D'}{\alpha_U' - \alpha_0'} < \frac{1 - \alpha_D}{\alpha_U - \alpha_0}$$

a) is trivial to demonstrate since the denominator increases and numerator decreases. Demonstrating b) requires more algebraic manipulation but can be shown to hold after all the appropriate substitutions are made.

Proposition 3

Given equations (6) – (9), we simply need to show that $\beta_U' - \beta_D' > \beta_U - \beta_D$ and $\beta_U' - \beta_0' > \beta_U - \beta_0$. This is trivial to show given the fact that we are in fact making the spread wider between the two types.

Proposition 4

Given equations (6) – (9), we need to show that $\frac{u - c}{(1 - \beta_0)u + \beta_0 c}$ is decreasing in c . This is the case since the derivative of this function is $\frac{-u}{[(1 - \beta_0)u + \beta_0 c]^2} < 0$.